A VERIFICATION SYSTEM FOR SHORT RANGE NAVY FORECASTS

BY

PAUL MARTIN WOLFF

A VERIFICATION SYSTEM FOR
SHORT RANGE NAVY FORECASTS

by
P. M. Wolff

A VERIFICATION SYSTEM FOR SHORT RANGE NAVY FORECASTS

by
Paul Martin Wolff
Lieutenant, United States Navy

Submitted in partial fulfillment
of the requirements
for the degree of
MASTER OF SCIENCE
IN AEROLOGY

United States Naval Postgraduate School
Monterey, California
1950

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

This work is accepted as fulfilling
the thesis requirements for the degree of

MASTER OF SCIENCE
IN AEROLOGY

from the
United States Naval Postgraduate School

PREFACE

This paper presents the author's attempt to develop a short range forecast verification for Navy use. It is intended to be a simple practical system yet based on sound meteorological and statistical principles.

Undertaken as the thesis requirement for the degree of Master of Science in Aerology, this paper was prepared at the United States Naval Postgraduate School, Monterey, California, during the academic year 1949-1950.

The author is deeply indebted to Professor W. D. Duthie, for the original suggestion of the subject and for his valuable assistance during the development. He also wishes to acknowledge the assistance rendered by Associate Professor A. Boyd Mewborn in the preparation of the verification tolerance scales.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

I. INTRODUCTION

The subject of forecast verification continues to be as controversial now as it was when the pioneering work in the field was done in the last century. The variety of opinion and counter opinion still exists, as there are no authoritative criteria for settling an argument of such a subjective nature. However, different systems will be required to fit the diverse uses of forecasts.

In this paper the requirements of a system which meets the particular needs of the United States Naval Aerological Service are set up. The various verification schemes proposed and used in the past are examined. The simple percentage of correct forecasts is rejected for several reasons. The need for some logical basis of comparison is then established.

Three classes of these bases for comparison are examined. They are climatology, pure chance and allied computations, such as skill score, and persistence. Persistence is established as the most logical practical basis for comparison for the Navy purposes in verification as set forth here. In fact, the first two schemes of comparison, climatology and chance, are positively rejected as unsuited for either theoretical or practical reasons or both. Climatological evidence is presented in support of the above contention.

Then with the basis for the verification system decided upon as a comparison with persistence, the details are developed in accordance with the laws of probability and statistics.

Before an objective verification can be attempted the terms in which the forecast is to be made must be rigidly defined. The exact form of the 36-hour forecast is therefore specified. Some features of the present Navy forecast form were retained but many innovations are made which should increase the preciseness and completeness with which the forecast is stated by the forecaster.

The development of the tolerance tables is then presented. The degree of correctness of the forecast is examined with each meteorological element considered separately. Differences with the present Navy system are evident here. Precipitation, cloud cover, and visibility are each considered separately.

The verification score is determined for each of eight meteorological elements forecast, viz. precipitation, average cloud cover, lowest ceiling, lowest visibility, surface wind direction, average hourly wind velocity, maximum single gust, and maximum or minimum temperature as appropriate. The score depends on two things, amount of change since the previous day and the degree of correctness of the forecast. The gradations of the tables in degree of correctness are matched with the limits of observerable or predictable accuracy. The logical basis for using the amount of change from the weather of the previous day, i.e., persistance, as a criterion of difficulty of a forecast, is established. The numerical score obtained by adding the scores for each forecast element, read in the appropriate tolerance table, has no percent significance. By taking a score of zero for a correctly forecast persistant occurrence, the superiority or deficiency of the forecast compared with persistance is automatically established.

The verification system is then tested on a series of forecasts for typical Navy stations, illustrating the similar total scores obtained from the verification of forecasts ranging widely in difficulty.

Finally, the proposed forecast form and verification system are compared with the system now in use by the Naval Aerological Service. The preciseness of the terms in which the forecast must be stated increases its value to the consumer and the time required for verification compares favorably with the present system.

II. REQUIREMENTS OF A VERIFICATION SYSTEM FOR NAVY USE

In order to devise a forecast verification system for Navy use, it is first desirable to enumerate the uses of verification of forecasts. From these uses, the particular qualities most desirable can be derived.

The principal uses of verification should be in studies by which forecasts could be improved. A study of errors may indicate consistent trends toward too radical or too conservative forecasts of particular weather elements.

Another widespread use of verification is in determining the maximum period for which forecasts are of value by establishing a significant minimum score. Also forecasters are frequently ranked in ability according to their accumulated forecasting average.

These uses in themselves may not appear to justify the time and energy required in verification. It is felt that accurate verification records will provide incentive for individual forecasters to increase their skill while the absence of verification will have a contrary effect.

These rather general uses of verification require several restrictive specifications when considered from a Navy viewpoint. Any verification system will inevitably be used to rank forecasters at one station and possibly among several different stations. The system must be based upon and developed by sound statistical principles. It is desirable also that the score obtained by verifying the forecast reflect only the skill of the forecaster. This requires a system that will give comparable scores at

one station for forecasts made in a variety of synoptic situations of widely varying difficulty. In addition the system should reflect only forecasting skill when applied in climates where tho weather is very changeable as well as those in which interdiurnal variation is slight.

III. TYPES OF VERIFICATION SYSTEMS AND THEIR HISTORICAL BACKGROUND

Completeness and continuity in the historical development of verification systems and their classification is largely made possible by the excellent survey of the literature by Muller [5]. This is especially true of the discussion of the contributions of European meteorologists prior to 1920.

The earliest verification systems were based on a simple computation of percentage of correct forecasts. This system is referred to as the percentage system. A forecast was judged to be a complete success (hit) or a complete failure (miss) according to fixed tolerances. For example, with a tolerance of 4 degrees, a forecast of any temperature 41 to 49 degrees inclusive would be judged a complete and equal success if the observed temperature were 45 degrees. Any other forecast temperature would be scored a failure. This is the percentage system with fixed tolerances. This rather naive method is still in use by some weather services although the meaninglessness of percentage of hits as a test of skill in forecasting was pointed out by Köppen [4] as early as 1906.

Modern users of this method commonly employ rating scales, assigning points according to the degree of success of the forecast. These ratings are then converted to percentages. The Naval Aerological Service is one of the very few still retaining simple percentage verification with fixed tolerances.

The second group of verification systems includes those in which the score is determined by the degree which the success of synoptic forecasts

differs from the success of some type of comparison forecast. This group is divided into types according to the kind of comparison forecast used. These types involve (a) elimination of pure chance; (b) comparison with some kind of persistence forecast; and (c) comparison with some kind of climatological forecast.

Köppen's original suggestion was a comparison with random forecasts. This would eliminate the portion of the forecast's success due to chance. The current use of skill score is another example of the elimination of chance successes. However, the use of skill score is limited to variates capable of representation in a tetra-chloric distribution, such as those which can be analyzed on a simple occurrence or non-occurrence basis. This type of analysis applied to thunderstorms or precipitation or other discrete weather phenomena is very effective.

A second type of comparison forecast was developed by Heidke [3]. He accepted Köppen's view that percentage of hits is an inadequate measure of forecasting skill. He proposed a system based on using the previous day's weather; i.e., persistence, as the comparison forecast. The actual verification score was obtained from complicated formulae. An interesting variation of this method was proposed by Dinies [2] for use in the German Weather Service. The observed weather of the particular day in the previous year was used for comparison.

More recently a third type of comparison forecasts has come into limited use. These systems use climatological records as a point of reference. Clayton's [1] extensive work and the method proposed by the Verification

Section of the Weather Division, Headquarters Army Air Force [7] are examples of this use of climatology.

In Muller's entire survey of the literature on verification, only one author included has published articles expressing the belief that forecasts should not be verified, Schmauss [6]. Although there is some modern support for this theory, certainly the great majority of meteorologists recognize the need for objective forecast verification.

IV. SELECTION OF TYPE OF VERIFICATION SYSTEM BEST FITTED TO NAVY USES

The first and simplest type of verification system is based on a calculation of the percentage of correct forecasts, with each forecast of an element being considered as a hit or a miss according to a fixed tolerance table. This system is presently used by the Naval Aerological Service but it has several serious faults when applied to one station and even more when applied to a variety of stations.

Forecasts vary in difficulty from day to day and from season to season at the same station. Thus forecast scores computed on a basis of percentage of hits will show wide variation from day to day, with seasonal trends probable, all independent of the skill of the forecaster.

The complexity and rapidity of weather changes which,in general increase with increasing latitude and vary with geographical location,will make impossible any comparison of forecasting scores from different locations such as the widespread Navy stations.

The tolerances used to determine whether a forecast is a success or failure in the percentage system are not suitable for stations with wide geographical differences. For example, with a tolerance of 4 degrees Fahrenheit in daily minimum temperature at a maritime low-latitude station, 100 percent hits might be possible on a repeated forecast of one particular temperature, while at an inland high-latitude station, great forecasting skill would be required to obtain a score of 75 percent hits on minimum temperature for the same month.

Thus a system employing percentage of correct hits seems extremely unlikely to fulfill the Navy requirements as outlined in Chapter II, even with considerable modification.

A system employing some type of comparison as a reference point and making some allowance for forecast difficulty seems necessary. The elimination of the successes achieved by pure chance is not adequate, since it does not change with changing forecast difficulty. Expectancies calculated on the basis of climatology seem most logical. Then the correct forecast of a rare event could be given higher weight than the correct forecast of a common event.

An attempt was made to set up such a system. Records were obtained from several typical Navy stations by recording the various weather elements as they appeared on synoptic charts for several months of all seasons of the year. The variations in average values and distribution of the various weather occurences were so great that tolerance scales for forecast score computed on the basis of long-term averages would have little meaning when applied to an individual season or month. For example, advection fog at Pensacola had a long-term average occurrence of eight 12-hour periods. The best estimate of the expectancy of fog at Pensacola for any 12-hour period would be 8/56. However, if this frequency were used in scoring verification of fog forecasts in February 1947 and February 1948 it would have little meaning, for in February 1947 fog occurred not at all and in February 1948 it occurred during twenty-three 12-hour periods.

Temperature records are cited in Table 1. These data show the variation of monthly mean temperatures about the long-term mean. This tabulation also shows that the variation of monthly means is so great that tolerance intervals based on long-term means will be too large to be practical.

INTER DIURNAL TEMPERATURE VARIATIONS IN DEGREES FAHRENHEIT

		-8	-8	-6	-4	-2	0	2	4	6	8	7 8
	1930	2		1	2	5	8	7	1	1	2	1
	1931	2		1	2	11	2	3	5	2	1	1
August	1933				2	5	16	11	3			
Kodiak	1934			1	4	5	8	7	5			
	1937	1	1	1	2	6	7	5	3	1	2	1
	1938	1	2	1	2	6	8	4	3	2		1
	Total	6	3	5	14	38	49	37	20	6	5	4
	1930		1	2	4	4	7	5	4	3		
	1931				7	5	6	6	4	2		
January	1933	2	1	3	1	5	6	4	3	3	1	1
Kodiak	1934	3		2	1	4	6	5	3	2	3	1
	1937	3	2	1	3	6	3	3	3	2	1	3
	1938	4		2	5	4	2	1	3	6	1	2
	Total	12	3	8	21	28	30	24	20	18	6	7

TABLE 1.

VARIATION OF MONTHLY MEAN TEMPERATURES ABOUT LONG TERM MEAN

A	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49
B	1	1	3	1	3	2	12	10	3	-2	-1	9	1	2	-6	-1	0	-1	3	-5	2	5	8	8
C	3	2	3	-8	-13	7	-2	4	11	3	1	-16	6	8	-1	5	-2	-4	2	6	6	0	M	-15
D	-1	3	0	3	-2	1	7	7	4	3	-1	10	2	4	-11	2	-1	2	0	0	1	4	M	6
E	1	2	1	-3	-4	4	-1	-2	4	1	4	-8	3	2	4	4	2	0	2	1	1	-2	M	-6

A Year
B New England
C Dakotas
D Central Gulf
E California

TABLE 2.

Frequency and probability of occurrence of the various weather elements computed after the forecast interval (month) is over would provide a logical basis for assessing difficulty. However, this is considered impractical. The forecaster should have prior knowledge of the verification tolerances. Aerological office routine makes a daily verification desirable.

The remaining basis for comparison of forecasts is persistence. This is less desirable theoretically than probabilities computed <u>post facto</u>, but is much more practical in application. Day-to-day persistence was chosen as being the most probable as well as most convenient standard. Persistence has the great advantage of universal application to any climate. The fact that the most probable weather for tomorrow is that which occurs today is supported by the data of Table 2.

V. DEVELOPMENT OF TOLERANCE TABLES AND SCORING SYSTEM

Having selected persistence as the basis for comparison, the details of the verification system remain to be decided upon. Several principles have been selected as guides in the development of the verification schemes for the various weather elements of the complete forecast. These are briefly discussed here along with the specific purposes they serve.

In the forecast and verification system presently used by the Naval Aerological Service, the method of presentation of state of the weather is considered particularly weak. A prescribed set of terms are available in this system and one of these must be chosen by the forecaster and entered in the appropriate space in the forecast form. These terms range from definite events, such as thundershower and showers to very indefinite terms, such as "mostly fair" and "threatening". These last two terms are defined as a certain range of cloud cover with a small amount of precipitation permitted. Fog (defined as visibility below a certain minimum for a certain duration) is another permissable term. Thus in this system, precipitation, cloudiness, and visibility are all considered in one forecast element, the state of the weather. The result can only be confusion and uncertainty in the minds of those who must use the forecast.

This difficulty will always be present when a word or short phrase is used to summarize the weather conditions of a period of time. A maximum of clarity can be obtained by stating the forecast for each weather element (precipitation, cloud cover, visibility, etc.) separately. The short summary spaces of the forecast form must be filled with specified terms whose meaning is clearly defined in the minds of the user as well as the forecaster.

In the preparation of the tolerance allowances of the proposed system it is assumed that the most probable occurrence for any forecast weather element for the following day will be the observed weather of the present day. It is further assumed that the distribution of tomorrow's weather values about today's is normal. This is supported by the data of Table 2. Although only temperature data are exhibited, shorter tests were run with other meteorological elements and the results were comparable.

The forecast elements are separated into three classes according to their type. Precipitation is analyzed on an occurrence or non-occurrence basis. Wind direction, cloud cover, and temperature are continuous variates with no variation in meteorological importance. That is, a Northwest wind direction is not considered any more or less important than any other wind direction. The third class of variates includes ceiling, visibility, and wind velocity which, although continuous, have varying importance. For example, with low visibilities an error of one mile is more important than the same error with 15 mile visibility. The change in importance was considered from a meteorological viewpoint. No attempt was made to assess the operational importance of various forecasts. This would vary widely according to the type of military operation for which the forecast is intended. For example, for high level photography an overcast at 5000 feet might be considered bad weather, while for surface ship operations it might be quite unimportant.

In all tolerance tables an attempt was made to verify to the same accuracy as the observation is made. When the observed maximum temperature is 65°, a forecast of 65° should get more credit than any other forecast, with the score decreasing rapidly with increasing error.

A fundamental question remaining is the scoring system. What ratio should be chosen between the credit allowed for a correct persistence forecast and the credit allowed for a correct change forecast?

The persistent weather of low latitude stations is reflected in the high forecasting scores obtained there with fixed-tolerance percentage verification. The rapidly changing weather of higher latitudes causes lower scores when forecasts are rated on the same system. This inequality is independent of forecast skill and should be removed.

After a study of percentage verification records from high and low latitude stations a ratio of six to ten was chosen. This value was computed from the average percentage scores at stations in both types of geographical location.

The tolerance tables employ these ratios. The decrease in credit with increasing error is computed from ordinates of the normal curve. Using a hypothetical forecast element as an example, the scoring for errors from 0-4 units from the normal ordinates would be:

Error in units	Score
0	10
1	9.6
2	8
3	5.4
4	3

In order to avoid fractions the scores are doubled and rounded off to the nearest whole number according to standard rounding procedure. The above example then becomes:

Error in units	Score
0	20
1	19
2	16
3	11
4	6

The abscissae in the tolerance table are fitted by the inverse normal ordinate according to the amount of change from previous day. In the hypothetical example, the zero error row might read across:

	Difference from yesterday in units				
Error	0	1	2	3	4
0	12	13	14	17	20

Then the complete table for this simplified illustration would be:

Error in units	Difference from yesterday in units				
	0	1	2	3	4
0	12	13	14	17	20
1	6	7	10	16	19
2	0	0	5	11	16
3	0	0	0	7	11
4	0	0	0	4	6

The point at which the forecast has zero value is arbitrarily selected for each table. Two-way tables such as the one above are used to verify wind direction, cloud cover, and temperature. The score is determined by the error in the forecast and the persistence of the element.

In the cases of those elements with varying natural importance a three-dimensional table is used. This requires two plane tables which determine forecast score from three variables: the meteorological importance of the forecast, the amount of persistence, and the error in the forecast.

The score for a correct persistence forecast is now translated from a value of twelve to zero. This serves two purposes. The correct persistence forecast automatically scores zero, making comparison possible without

computation. Also it removes any possibility of the final score being confused with a percentage score on the basis of 100 percent for a correct forecast. Several objective verification systems in the past, notably those of Heidke and Clayton, have been considerably criticized by other forecasters solely because the numerical score computed by their system was small compared to 100 percent.

After this translation of reference point, the example above would appear as follows:

	Difference from yesterday				
Error	0	1	2	3	4
0	0	1	2	5	8
1	-6	-5	-2	4	7
2	-12	-12	-7	-1	4
3	-12	-12	-12	-5	-1
4	-12	-12	-12	-8	-6

Special details of the scoring system as applied to the individual forecast elements are discussed here.

1. Precipitation. The type or intensity of precipitation is of little importance for Naval uses when the effect on ceiling and visibility is adequately forecast. Quantitative forecasting of precipitation is beyond the present precision of the science. For these reasons precipitation forecasts are analyzed on a simple occurrence or non-occurrence basis.

The only modification introduced is to cover cases in which only very small amounts of precipitation occurred. In such cases a forecast of no precipitation would be given half credit. Amounts up to and including .02 inches are defined as slight precipitation for this special modification.

2. Cloud Cover. No special details.

3. Ceiling. The meteorological importance is assigned roughly in accordance with the Civil Aeronautics authority requirements for instrument and contact flights. The error is assessed in 500 ft. units.

4. Wind Direction. An eight-point compass was considered adequate for general meteorological uses. Any forecast with an error of more than two points was considered worthless.

5. Wind Velocity. Although average velocity is computed to the nearest knot, the use of a two-knot interval in computing error was necessary for brevity. The increased accuracy attained by the larger table necessary to include one-knot intervals is not justified. The change of scores would be significant in only one or two places in the entire table.

6. Maximum Gust. The present verification system verifies maximum hourly velocity. This value has little meaning, so maximum single gust was substituted. It is forecast and verified to five-knot intervals.

7. Visibility. The meteorological importance was again decided in accordance with requirements for CAA Closed, Instrument, and Contact flight regulations.

8. Maximum and Minimum Temperatures. Verification tables have two-degree intervals for brevity although temperature is observed to the nearest whole degree. The small errors occurring where the slope of the normal curve is great and the difference in scores between spaces in the table relatively large are discussed under wind velocity above.

VI. PROPOSED FORECAST FORM

Before any objective verification can be attempted, the forecast must be stated in precise meaningful terms. The specifications for each forecast element require a complete concise statement of the expected weather with no chance for vague terms or "hedges". The practice of hedging in forecasting is old and widespread. It is largely responsible for the popular misunderstandings of the possible achievements and also the limitations of the science. In this proposed verification system the maximum score is attained only by a very accurate forecast.

With forecast systems that permit indefinite complicated terms such as "mostly fair", there is a strong tendency for forecasters to attempt to include in their forecast all possible synoptic developments. When fixed tolerance systems are not clearly thought out there may exist favored numerical forecasts. For example, in the system in current Naval use the tolerances change in too large intervals. Thus with an observed average wind velocity of ten knots the allowance for a success is plus or minus two knots while the allowance for a hit with an observed average velocity of eleven knots jumps to four knots. This favors forecasts of nine or eleven knots and makes the use of ten knots penalize the forecaster. Numerous other cases of this inconsistency exist in the current system.

In this proposed forecast and verification system, the forecast is required to be stated in simple precise fashion and the verification scoring does not encourage any attempt to hedge.

The present box system of NavAer 447(a), containing a brief statement of the values to be verified is considered desirable. These boxes contain the forecast in explicit unambiguous terms, readily available for verification. These boxes are arranged in a vertical column with the detailed forecast opposite them.

The forecast form suggested here is for a 12-hour forecast interval. Three of these forms would ordinarily comprise a normal forenoon forecast. However, two forms for day and following night could be used for a preliminary (early morning) forecast and the length of the forecast period could be extended by adding more of the basic 12-hour forms.

Figure 1 is an example of the 12-hour forecast form. Figure 2 is the same form containing a sample forecast. The specifications for the forecast are set forth below.

Precipitation. Box: yes or no as appropriate. Detail: if box contains yes specify type or types of precipitation, time of beginning and ending if within forecast period, intensity and changes in intensity, and amount of precipitation expected in period. The special classification, light precipitation, is intended for use only in forecast verification.

Sky condition. Boxes: state average cloud cover in tenths for period in upper box. In lower box specify lowest ceiling expected to occur for two successive observations. Detail: specify chronological variation of cloud cover. State maximum and minimum number of tenths expected and give times of occurrence. Specify chronological variation of ceilings, including highest and lowest ceilings expected and times of occurrence. Include statement of turbulence and icing when applicable.

Visibility. Box: lowest visibility in miles expected to occur for two successive observations. Detail: specify chronological variation of visibility.

Surface wind. Boxes: top box for prevailing wind direction during period. Middle box for average hourly velocity in knots. Lower box for maximum single gust in knots. Detail: specify all significant changes in direction, velocity or gustiness. Include highest and lowest hourly averages expected with times of occurrence.

Temperature. Box: maximum or minimum temperature in degrees Fahrenheit as applicable. Detail: specify variation of temperature by giving forecast temperature for 4-hour intervals during period. Include time of maximum or minimum, and where applicable, the time at which the temperature is expected to reach freezing point (32°F) whether rising or falling.

Winds aloft. Not verified. Forecast for two times to coincide with local pilot balloon observations during forecast period. Are to be forecast for 6 levels to be chosen by the forecaster according to his operational commitments.

Fig. 1 PROPOSED FORECAST FORM

Local Time Forecast Effective

Precip Yes No	Precipitation

Ave. Cloud Cover Tenths	Sky Condition
Lowest Ceiling Hundreds of feet	

Lowest Visibility Miles	Visibility

Wind Dir.	Surface Wind
Ave. Wind Vel. Knots	
Max. Gust Knots	

Max. Min Temperature °F	Temperature

Winds Aloft

Time				Time			
Level 1	Dir-Vel	Level 4	Dir-Vel	Level 1	Dir-Vel	Level 4	Dir-Vel
Level 2	Dir-Vel	Level 5	Dir-Vel	Level 2	Dir-Vel	Level 5	Dir-Vel
Level 3	Dir-Vel	Level 6	Dir-Vel	Level 3	Dir-Vel	Level 6	Dir-Vel

Fig. 2 EXAMPLE OF TWELVE HOUR FORECAST ON PROPOSED FORM

Forecast Effective 1800-0600 Local

Precip Yes	**Precipitation** Light showers accompanying weak cold frontal passage at 2200 local. Showers occurring 2000 local to 2400 local. Precipitation .10 inch expected.
Ave. Cloud Cover 8 tenths Lowest Ceiling 2000 feet	**Sky Condition** Five tenths small cumulus clouds based at 3000 ft. at 1800 local increasing to overcast with showers. Ceilings lowering to 2000 ft. in showers. Tops 4500 ft. increasing to 7000 ft. in showers. Clouds flattening and decreasing to five tenths strato-cumulus base 2500 ft. tops 3500 ft. at 0600. Light to moderate icing in clouds above 4000 ft. after 2400. Light to moderate turbulence during frontal passage.
Lowest Visibility 3 Miles	**Visibility** Visibility 7 miles hazy at 1800, decreasing to 3 miles in showers 2000-2400 and increasing to 15 miles at 0600.
Wind Dir. NW Ave. Wind Vel. 12 Max. Gust 30	**Surface Wind** Surface winds SW 10 - 15 knots with gust to 20 knots veering to NW with gusts to 30 knots at 2200 local and decreasing slowly to NW 8-12 knots at 0600. Highest hour 15 knots 2100-2200. Lowest hour 8 knots 0500-0600. Maximum gust expected about 2200 local.
Minimum 37	**Temperature** 1800 local 49 degrees 2200 local 46 degrees 0200 local 43 degrees 0600 local 37 degrees minimum

2100 local		winds		aloft		0300 local	
2000'	210-30	15000'	240-45	2000'	310-22	20000'	245-40
5000'	220-35	20000'	250-50	5000'	280-25	30000'	255-50
10000'	235-35	30000'	260-65	10000'	240-30	40000'	270-85

VII. DETAIL OF VERIFICATION

1. Precipitation. Enter Table 3 if forecast is correct and record score.

2. Cloud Cover. An average cloud cover in tenths is recorded for each hour. A maximum of ten tenths is permitted. The average of these values is used for verification. Enter Table 4 with the error of the forecast in tenths and the change from yesterday in tenths. Record this score.

3. Wind Direction. Record a direction to the nearest of eight points from the record of the selsyn recorder. Use direction which prevails for greatest number of hours. Enter Table 5 with error of forecast in points and the difference in direction from yesterday in points.

4. Maximum or Minimum Temperature. Obtain maximum or minimum temperature to nearest whole degree from hourly and check observations and/or maximum or minimum thermometer or thermograph if accurate. Enter Table 6 with error in degrees and difference from yesterday in degrees. Record score.

5. Average Hourly Wind Velocity. From single or multiple register record total number of knots passing anemometer during forecast period and obtain average hourly velocity to nearest whole number. Enter Table 7 with this velocity and yesterday's average velocity and obtain indicator letter. Enter Table 8 with indicator letter and error in knots and record score.

6. Maximum Gust. From Selsyn Recorder record of bridled anomometer obtain maximum single gust during period. Enter Table 7 with this velocity and yesterday's average velocity and obtain indicator letter. Enter Table 8 with indicator letter and error in knots and record score.

7. Ceiling. From airways record, special, and check observations obtain lowest ceiling occurring for two consecutive observations during the forecast period. Enter Table 7 with this ceiling and yesterday's lowest ceiling and obtain indicator letter. Enter Table 9 with indicator letter and error in nearest 500-foot units. Record score.

8. Visibility. From record, special, and check observations obtain lowest visibility occurring for two consecutive observations during the forecast period. Enter Table 7 with this visibility and yesterday's minimum visibility and obtain indicator letter and error in miles and record score.

Add these eight element scores to obtain final score.

As an example the sample forecast exhibited in Figure 2 is verified below.

Element	Previous Night	Forecast	Observed	Score
precipitation	none	yes	yes	16
average cloud cover	5 tenths	8 tenths	10 tenths	- 8
lowest ceiling	above 10,000 ft.	2000 ft.	1200 ft.	- 2
wind direction	SW	NW	SW	- 8
ave. velocity	7	12	14	- 1
max. gust	12	30	28	1
min. temperature	43	37	42	-12
			Total	- 4

It is intended that the scores for each 12-hour period be kept separate rather than be added together or averaged in any way. The increasing difficulty of periods farther away from the forecast time makes averaging of two or more of these periods detract from the meaning of the scores attained.

PRECIPITATION

	Score
Correct Change Forecast	16
Correct no Change Forecast	0
Incorrect Change Forecast	-
Precipitation less than .02 inch	-4
Incorrect no Change Forecast	-
Precipitation less than .02 inch	-12
Other Incorrect Forecasts	-24

TABLE 3.

CLOUD COVER

Change from Previous Day in Tenths

		0	1	2	3	4	5	6	7	8	9	10
	0	0	0	2	2	4	4	6	8	10	14	16
	1	-2	-2	0	0	2	2	4	6	8	12	14
Error in Tenths	2	-12	-10	-10	-8	-8	-6	-4	0	4	8	10
	3	-18	-18	-16	-16	-16	-14	-12	-8	-4	0	2
	4	-24	-24	-24	-24	-24	-18	-16	-14	-10	-6	-4
	5	-24	-24	-24	-24	-24	-24	-24	-18	-16	-12	-10
	6	-24	-24	-24	-24	-24	-24	-24	-24	-24	-18	-16

TABLE 4.

TEMPERATURE

Change from Previous Day in Degrees F.

		0-1	2-3	4-5	6-7	8-9	10-11	12-13	14-15	16-17	18
	0-1	0	1	1	2	2	3	4	5	7	8
	2-3	-6	-5	-3	-2	-1	1	3	4	6	7
Error in Degrees	4-5	-12	-12	-8	-7	-6	-3	-2	-1	1	4
	6-7	-12	-12	-12	-12	-12	-8	-7	-5	-4	-1
	8-9	-12	-12	-12	-12	-12	-12	-12	-8	-7	-6

TABLE 5.

WIND DIRECTION

Change from Previous Day in Points

		0	1	2	3	4
	0	0	1	2	5	8
Error in Points	-1	-4	-3	-2	1	5
	2	-8	-8	-7	-5	-2

TABLE 6.

[illegible]

[illegible]

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

[illegible]

[illegible]

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

INDICATOR LETTERS

		Visibility in Miles	Gust in Knots	Average Velocity Knots	Ceiling hundreds of feet
	1	0-1	60	45	0-5
Class	2	2-3	36-60	26-45	6-15
	3	4-10	26-35	16-25	16-50
	4	10	0-25	0-15	50

Class Difference from Previous Day

		0	1	2	3
	1	D	C	B	A
Class Today	2	G	F	E	
	3	J	I	H	
	4	N	M	L	K

TABLE 7.

VERIFICATION SCORING FOR WIND VELOCITY

	Maximum Gust					Average Velocity								
	Error in Knots					Error in Knots								
	0	5	10	15	20	0-1	2-3	4-5	6-7	8-9	10-11	12-13	14-15	16-17
A	8	5	0	-4	-7	8	8	7	6	4	0	-4	-7	-10
B	6	4	-1	-5	-8	6	6	5	4	2	-1	-5	-8	-10
C	4	2	-1	-5	-8	4	4	3	2	0	-2	-6	-8	-10
D	3	1	-2	-6	-8	3	2	1	-1	-3	-5	-8	-10	-12
E	5	2	-1	-7	-9	5	4	2	-1	-4	-6	-9	-12	-12
F	3	1	-3	-8	-12	3	2	1	-2	-5	-7	-9	-12	-12
G	2	0	-4	-8	-12	2	0	-3	-7	-9	-12	-12	-12	-12
H	4	0	-6	-12	-12	4	3	-2	-6	-9	-12	-12	-12	-12
I	1	-2	-7	-12	-12	1	0	-4	-7	-9	-12	-12	-12	-12
J	1	-2	-7	-12	-12	1	0	-4	-7	-9	-12	-12	-12	-12
K	3	-4	-12	-12	-12	3	1	-3	-8	-12	-12	-12	-12	-12
L	1	-5	-12	-12	-12	1	0	-4	-8	-12	-12	-12	-12	-12
M	0	-6	-12	-12	-12	0	-1	-5	-9	-12	-12	-12	-12	-12
N	0	-7	-12	-12	-12	0	-1	-6	-9	-12	-12	-12	-12	-12

TABLE 8.

[illegible]

[illegible] [illegible]

[illegible] [illegible]

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

CEILING AND VISIBILITY

	Lowest Ceiling							Lowest Visibility					
	Error in 500 feet Units							Error in Miles					
	0	1	2	3	4	5	6	0	1	2	3	4	5
A	8	5	0	-6	-12	-12	-12	8	6	-1	-7	-12	-12
B	6	4	-1	-7	-12	-12	-12	6	5	-2	-8	-12	-12
C	4	2	-3	-7	-12	-12	-12	4	3	-3	-8	-12	-12
D	3	1	-3	-8	-12	-12	-12	3	2	-4	-8	-12	-12
E	5	3	-2	-7	-9	-12	-12	5	3	-1	-5	-9	-12
F	3	1	-3	-8	-10	-12	-12	3	2	-2	-6	-9	-12
G	2	0	-4	-8	-10	-12	-12	2	0	-3	-7	-9	-12
H	4	3	0	-3	-7	-9	-12	4	3	0	-3	-7	-9
I	1	0	-2	-4	-7	-9	-12	1	0	-2	-4	-7	-9
J	1	0	-2	-5	-8	-10	-12	1	0	-2	-5	-8	-10
K	3	2	1	-2	-5	-7	-9	3	2	-1	-4	-7	-9
L	1	0	-1	-3	-6	-8	-9	1	0	-2	-4	-7	-9
M	0	-1	-2	-4	-6	-8	-10	0	-1	-3	-5	-8	-10
N	0	-1	-2	-4	-6	-8	-10	0	-1	-3	-6	-8	-10

TABLE 9.

[illegible]

[illegible] [illegible]

[illegible] [illegible]

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

VIII. CONCLUSION

As was stated earlier the final score computed as above has no percentage connotation. The highest score attainable is 80. From the amount of change from previous day required it is very unlikely that a score higher than 40 would be possible. If the weather were exactly the same as the previous day and were correctly predicted the computed score would be zero.

Several writers have expressed the belief that blind persistence, forecasting no change day after day, would give about 50 percent successes scored by fixed tolerance verification. This type of pure persistence was tested with the scoring proposed here and the average score was -55.

In another test a small group of forecasts were prepared and verified. for a selected group of naval stations involving a wide variety of geographical locations. In this test five forecasts were made for the following daytime period 0600-1800 and verified with the proposed tolerance tables. The average scores were: Boston -40, Pensacola -35, Coco Solo -42, Honolulu -42, San Francisco -39, Kodiak -36. The sample is certainly small and the scores may well reflect the personal forecasting experience of the author, but it is believed that they are fairly representative. The average score on these forecasts was -39. The uniformity of scores from high and low latitude indicates that the system has compensated fairly well for varying difficulty of forecasts. If any bias still exists, it is probably in favor of the higher latitude stations with their more difficult, changing weather.

The final determination of the average attainable score and how well the system fulfills its primary purpose of measuring skill in forecasting will await wider use and tests in a much larger number of cases.

The proposed forecast form is much better suited to Naval uses than the present form because of the separation of forecast elements and the completeness with which they must be forecast in unequivocal terms. This forecast form will require that the meteorologist have a clear picture of the expected weather in mind when making the forecast, as the specifications call for the values of most weather elements to be reduced to numbers stating highest, lowest, and average values expected and to state as closely as possible the time of occurrence. Definiteness as to event and time of occurrence diminish as the period of time of the forecast increases. However, a forecaster should be able to specify the weather within close limits for at least three of the proposed 12-hour periods.

Another use of this verification system is in the evaluation of the so-called objective or "mechanical" forecasts. As long as these forecasts are used in definite forecasts of an occurrence or non-occurrence type there would be no significant advantage in using this type of verification. The type of forecasts made according to inflexible procedures and from predictants whose connection with the variable forecast is not known should be confined to forecasts of a yes - no type.

When mechanical techniques are used to predict the value of continuous variates, such as wind, sky condition or temperature, any defects will be obvious when the forecasts are verified by this proposed system. The low scores occurring as a result of the large errors made by a mechanical system will overbalance their successes, especially if the system does not contain all significant predictants which affect the weather occurring.

The chief advantage of the verification system over the one presently used is that the final score more nearly reflects only forecasting skill. The scores obtained on successive days at the same stations and those obtained at different stations are comparable. The verification system offers incentive to the forecaster at both tropical and high latitude stations, requiring definite effort at both places to obtain high scores. The verification system leads the forecaster to state the forecast as precisely as possible, where the present system encourages hedging with indefinite terms and uses tolerance tables which favor certain numerical forecasts.

After more use or tests of the proposed verification system, it may be desirable to alter the degree of difficulty in some of the tolerance tables by increasing or decreasing certain values. Further tests may show the desirability of a further shift of the zero point from the value used here. A possible new reference point is the average score attained by persistence forecasts instead of the score of a correct persistence forecast.

BIBLIOGRAPHY

1. Clayton, H. H. Verification of Weather Forecasts. American Meteorological Journal. 6: 211-219, 1889.

2. Dinies, E. Vorkersageprufungen. Deutscher Flugwetterdienst: Reichsant fur Wetterdienst, Sonderband 5, Teil 2 XII-XIV, 1936.

3. Heidke, P. Ergebnisse einer objektiven Prüfung von Wettervorhersagen, Meteorologische Zeitschrift 52: 487-490, 1935.

4. Köppen, W. Wie erkennt man Blindlingsprognosen. Meteorologische Zeitschrift. Hann - Band, 347-356, 1906.

5. Muller, R. H. Verification of Short Range Forecasts (A Survey of the Literature). American Meteorological Society Bulletin. 25: 18-27, 47-53, 88-95. January, February, March, 1944.

6. Schmauss, A. Die Treffsicherhert der Prognosen. Das Wetter 28: 68-71, 167-68, 1911.

7. U. S. Army Air Forces Headquarters Weather Information Branch. Report No. 602. Short Range Forecast Verification Program. Washington, D. C. November, 1943.